Studying chromatin accessibility

Master in Omics Data Analysis

Beatrice Borsari
University of Vic, Vic
Computational Biology of RNA Processing, CRG, Barcelona

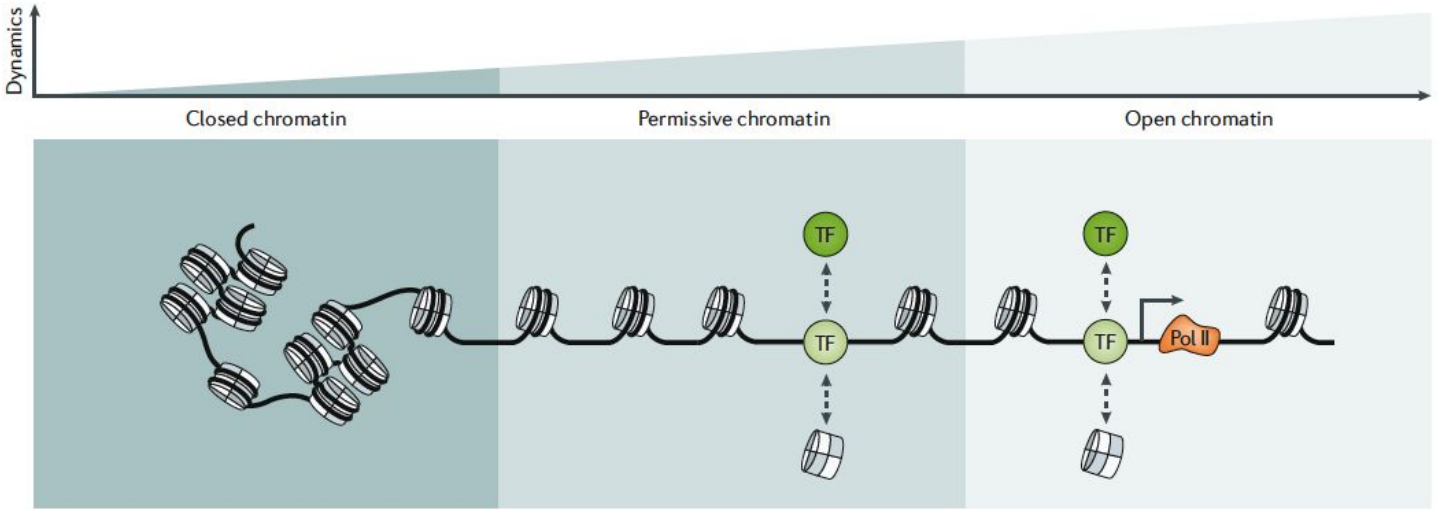
UVIC UNIVERSITAT DE VIC
UNIVERSITAT CENTRAL DE CATALUNYA

Master in Omics Data Analysis

CRG Centre for Genomic Regulation

Table of contents

Heterochromatin and euchromatin



heterochromatin

1. tightly packed, high nucleosome occupancy
2. low nucleosome turnover
3. little or no transcriptional activity
4. constitutive vs. facultative

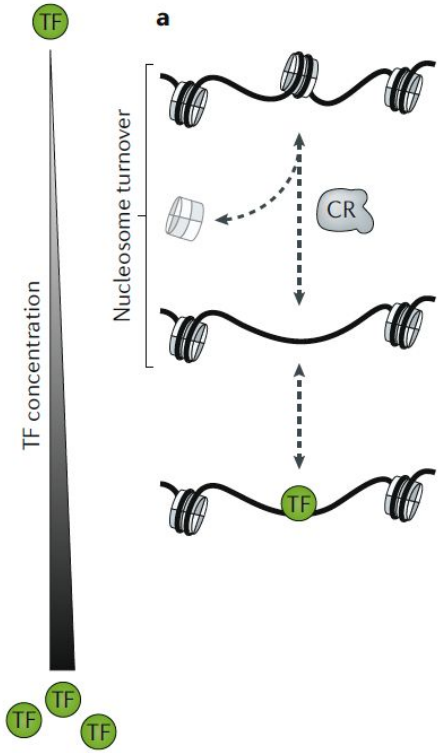
euchromatin

1. loosely packed, low nucleosome occupancy
2. high nucleosome turnover
3. can be bound by chromatin-binding factors
4. can be transcriptionally active

Klemm et al., 2019

Chromatin accessibility: fact sheet

- is the extent to which macromolecules inside the nucleus can physically contact chromatinized DNA
- is non-uniform across the genome
- the accessible genome comprises
  - ~2-3% of total DNA sequence
  - ~90% of regions bound by TFs
- depends on the distribution of nucleosomes and other chromatin-binding factors
- can alter transcription factor (TF) binding
- can change in response to external stimuli and developmental cues
- ATP-dependent chromatin remodelling complexes can move, eject or restructure nucleosomes

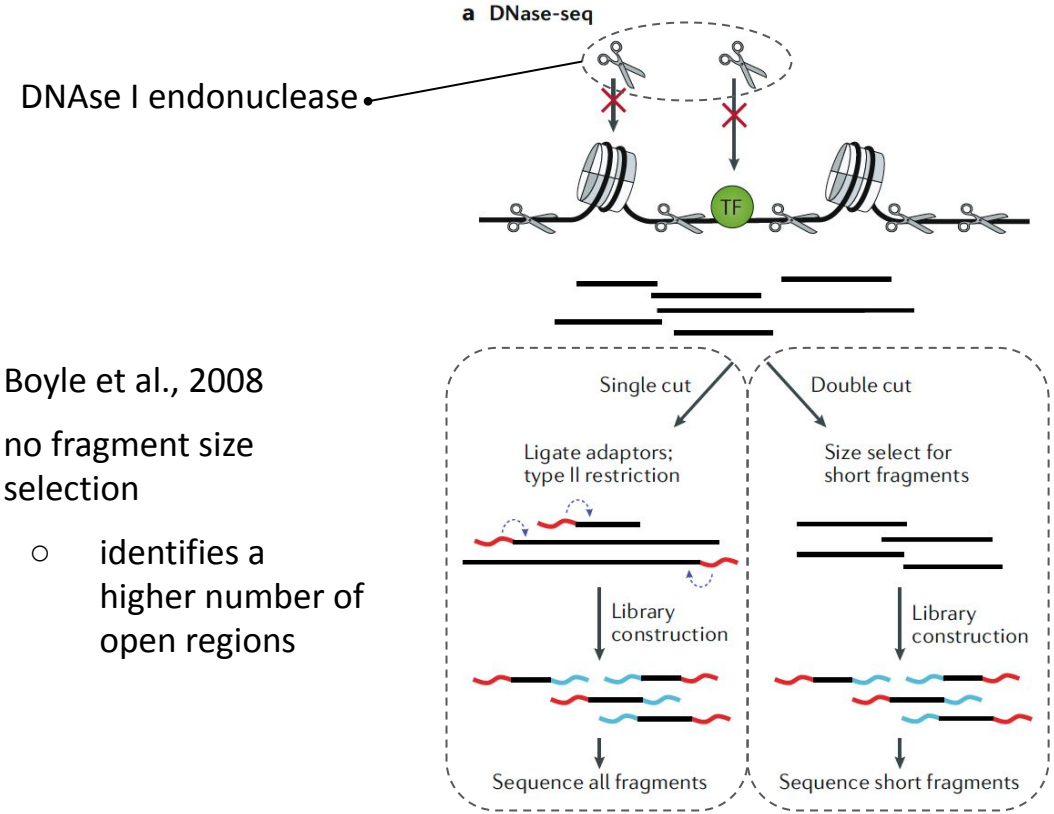


Klemm et al., 2019

Experimental techniques to study chromatin accessibility

- **DNase-seq** (Crawford et al., 2006; Sabo et al., 2006)
- **FAIRE-seq** (Giresi et al., 2007)
- **ATAC-seq** (Buenrostro et al., 2013; Corces et al., 2017)
- **MNase-seq** (Mieczkowski et al., 2016; Mueller et al., 2017)
- NOMe-seq (Kelly et al., 2012)

DNase-seq



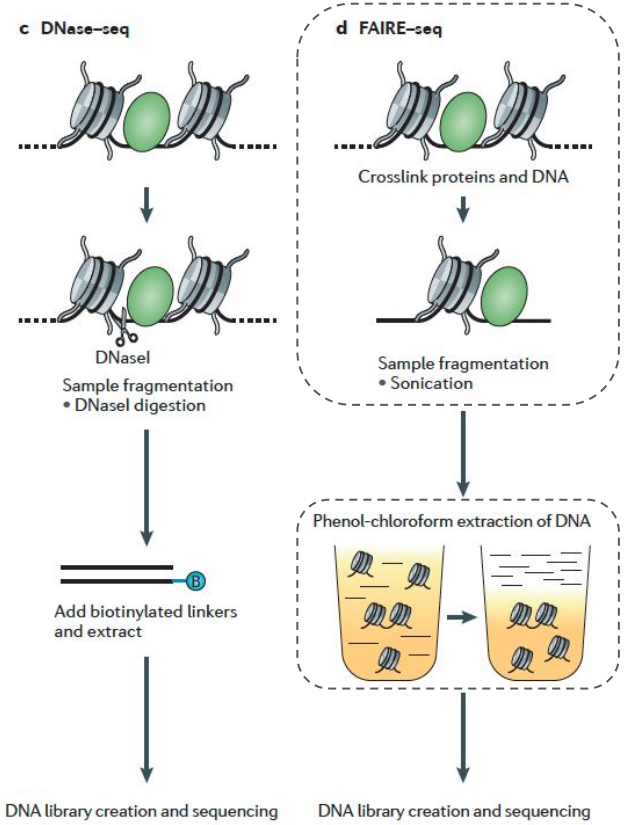
≥ 80% of open regions are far away from promoters → distal enhancers

- Boyle et al., 2008
- no fragment size selection
  - identifies a higher number of open regions

- Hesselberth et al., 2009
- selects short fragments ~ fragments originated by very close cleavage sites
  - less chances to sample fragments coming from inaccessible regions

Klemm et al., 2019

FAIRE-seq (Formaldehyde-Assisted Identification of Regulatory Elements)



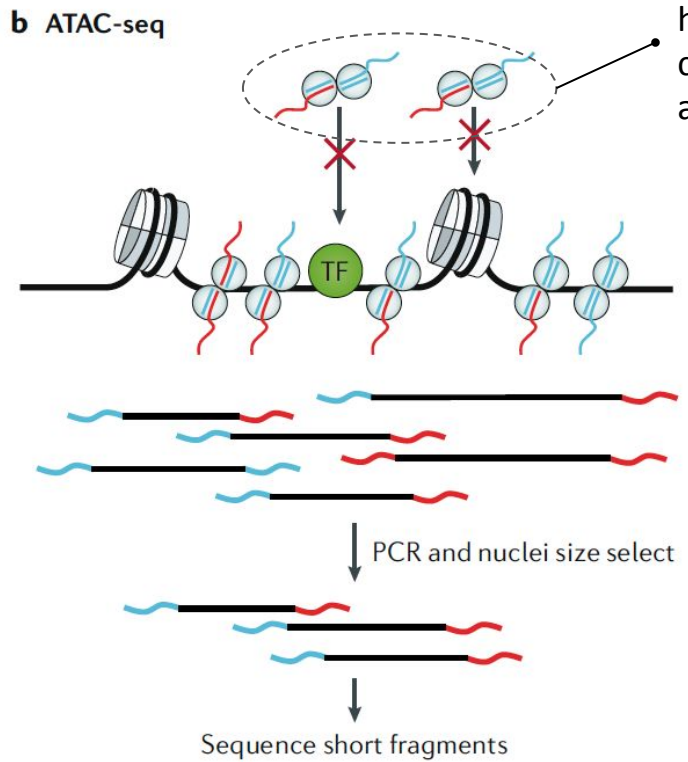
Similar to ChIP-seq

The nucleosome-depleted fraction of DNA is preferentially segregated to the aqueous phase.

- Both DNase-seq and FAIRE-seq show good correspondence with ChIP-seq data of different TFs
- Still, there are some differences:
  - FOXA1, FOXA3 and GATA1 better identified by FAIRE–seq
  - ZNF263 and CTCF better identified by DNase–seq
- DNase-seq-specific regions enriched at promoter regions, high overlap with H3K4me3, H3K9ac
- FAIRE–seq-specific regions enriched in introns, exons, intergenic regions; high overlap with H3K4me1

Furey, 2012

ATAC-seq (Assay for Transposase-Accessible Chromatin)

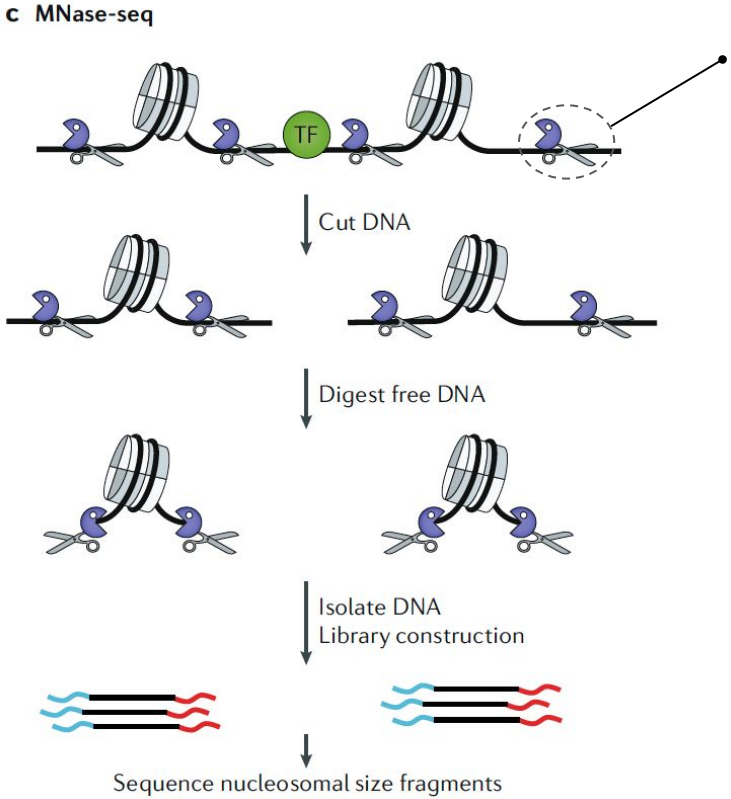


hyperactive transposase (Tn5): cleavage + binding of Illumina adapters

- similar to double-cut DNase-seq protocol: amplifies double-cleavage events
- very high correlation with double-cut ($r > 0.8$) & single-cut ($r > 0.75$) DNase-seq
- low amount of starting material (~ 500 cells)
- 2h < with 10,000-20,000 cells
- single-cell assays widely used: how do they work?

Klemm et al., 2019

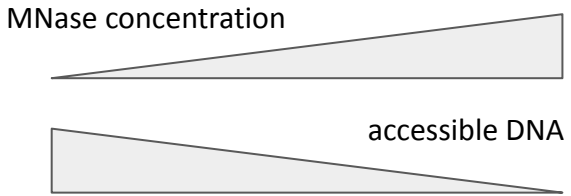
MNase-seq (Micrococcal Nuclease)



MNase:
1. endonuclease activity towards inter-nucleosomal DNA
2. exonuclease activity towards cleaved (free) DNA

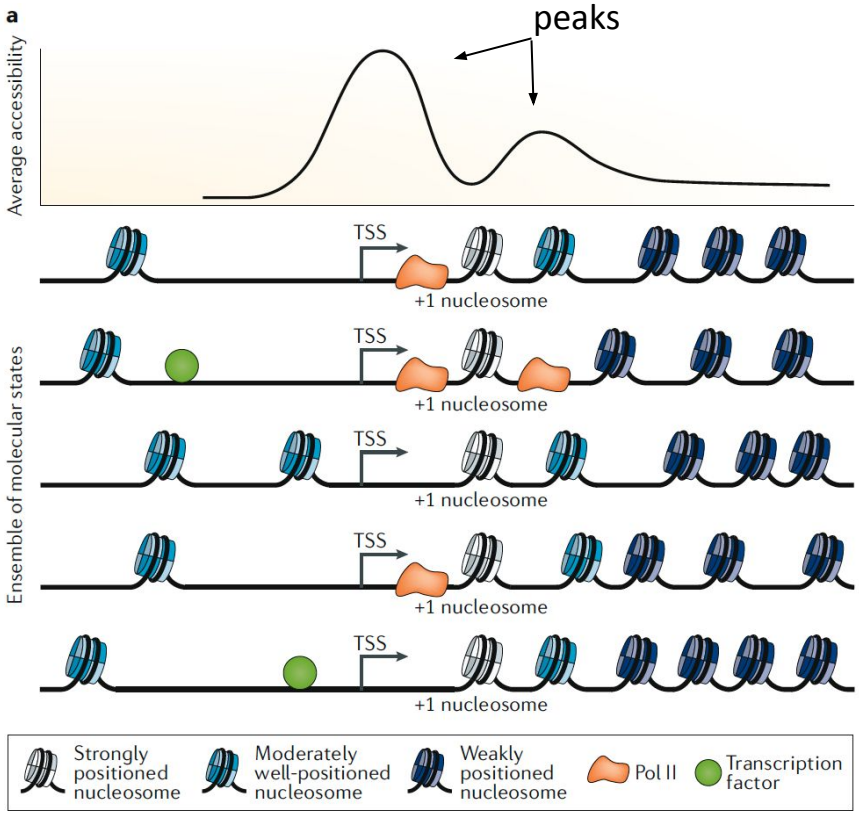
Differently from the previous techniques, it maps nucleosome occupancy and positioning

- the sequenced fragments correspond to nucleosomal DNA



Klemm et al., 2019

Interpreting chromatin accessibility profiles



cell 1

cell 2

...

cell *i*

cell *n*

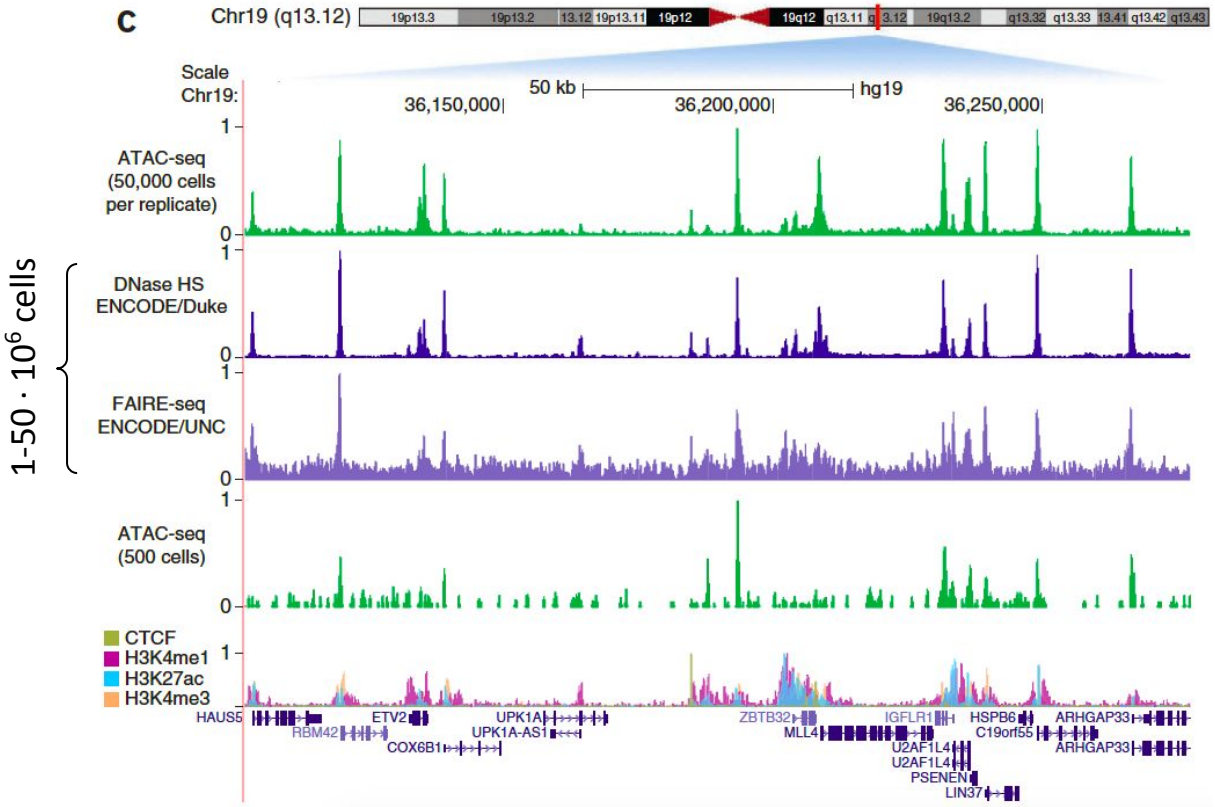
Chromatin remodelling is dynamic and unsynchronized

Heterogeneous cell population

We observe an average of different molecular states

Klemm et al., 2019

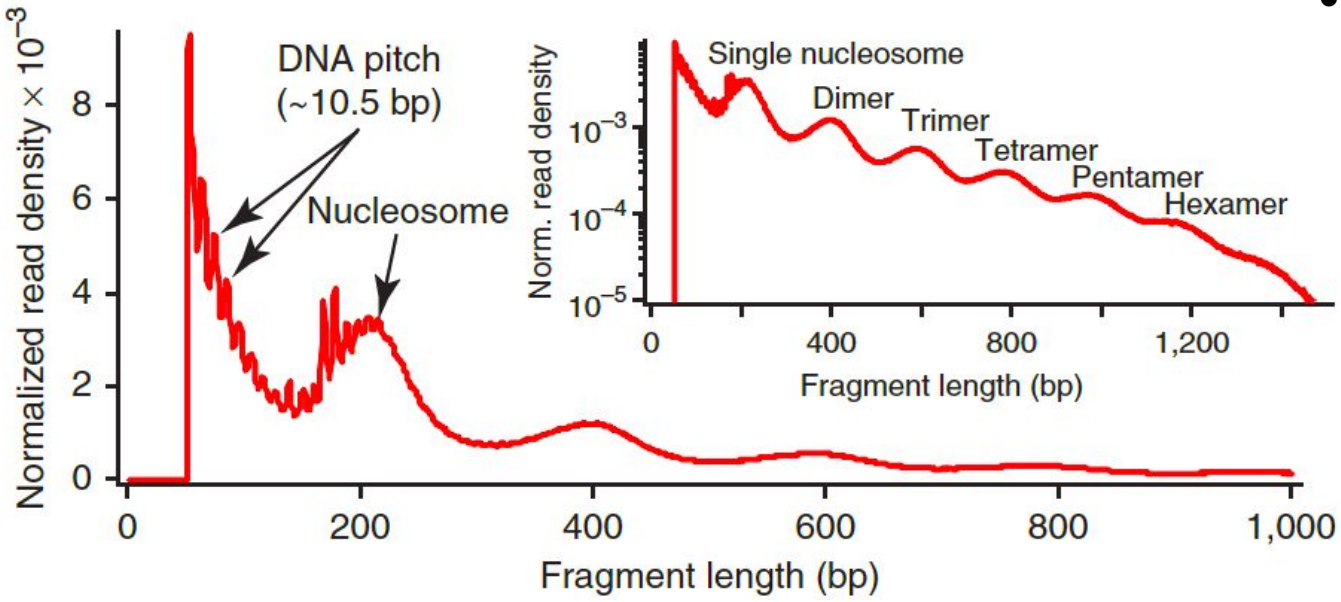
Comparing different -seq techniques to access chromatin accessibility



- ATAC-seq can be performed with lower amount of starting material
- Has a FC signal comparable to other techniques

Buenrostro et al., 2013

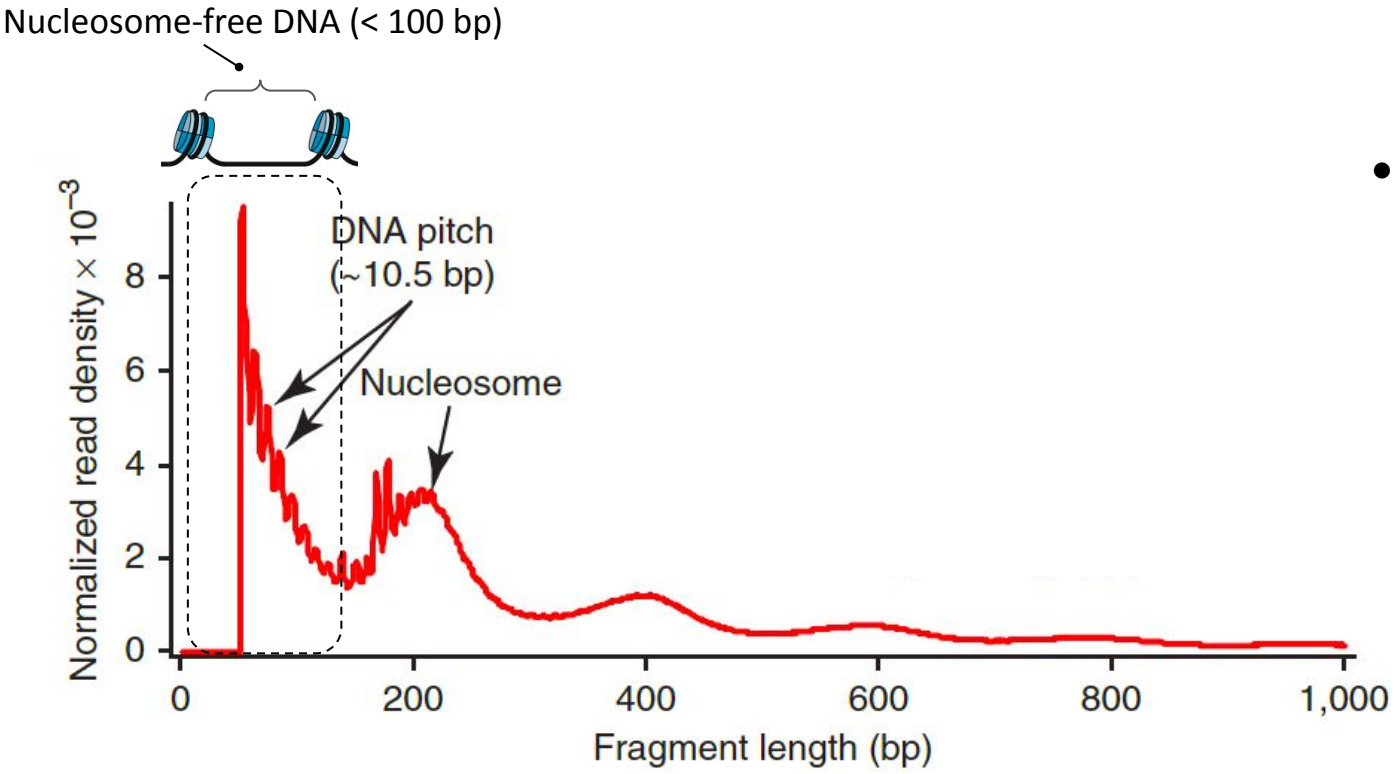
ATAC-seq proves the periodicity of nucleosome positioning



- periodicity of ~200 bp (~ nucleosomal DNA)

Buenrostro et al., 2013

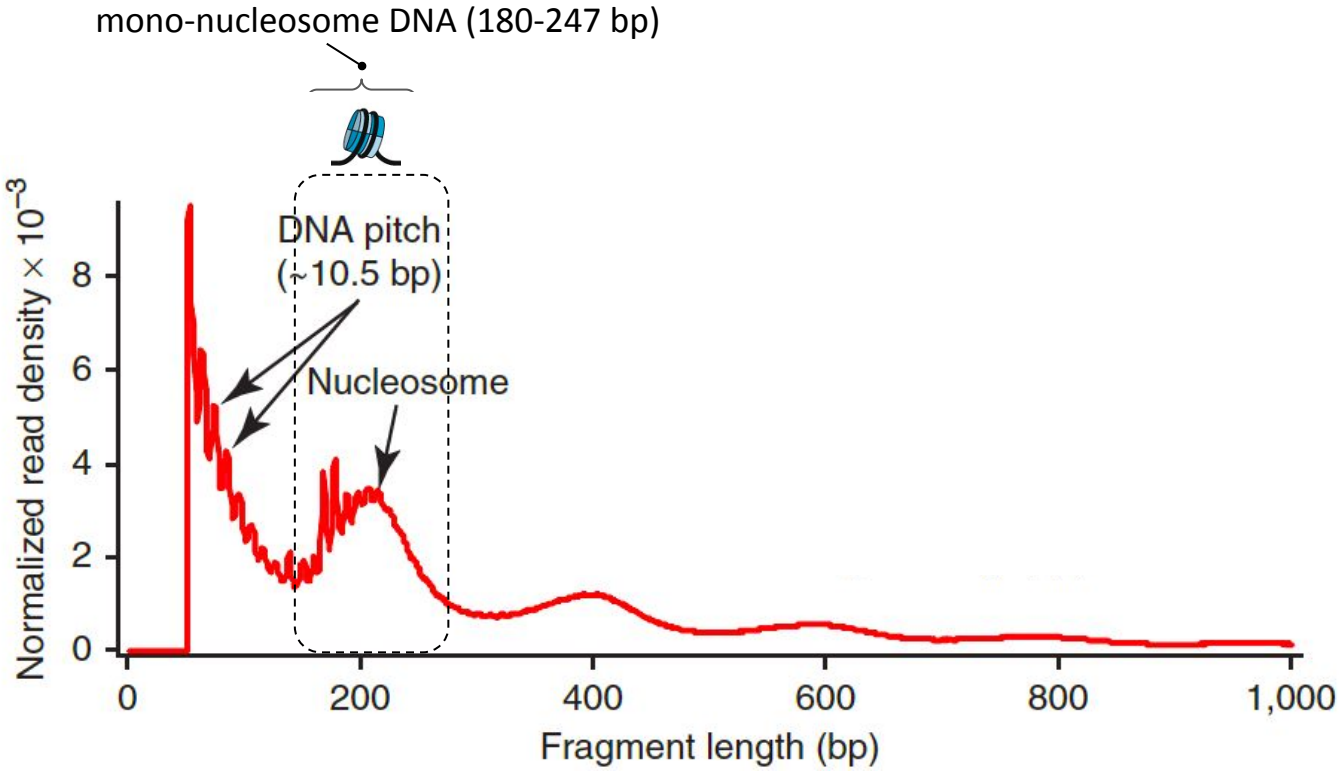
ATAC-seq proves the periodicity of nucleosome positioning



- periodicity of ~200 bp (~ nucleosomal DNA)

Buenrostro et al., 2013

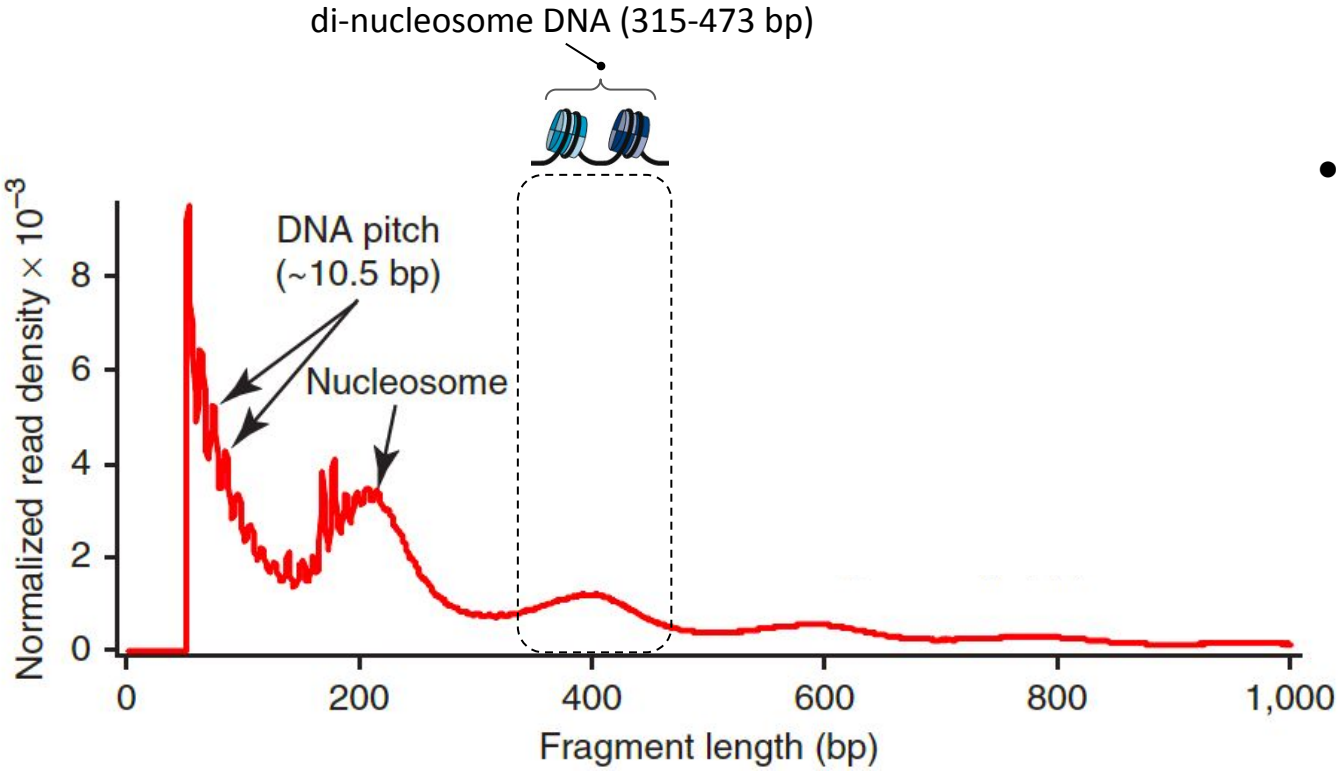
ATAC-seq proves the periodicity of nucleosome positioning



periodicity of ~200 bp
(~ nucleosomal DNA)

Buenrostro et al., 2013

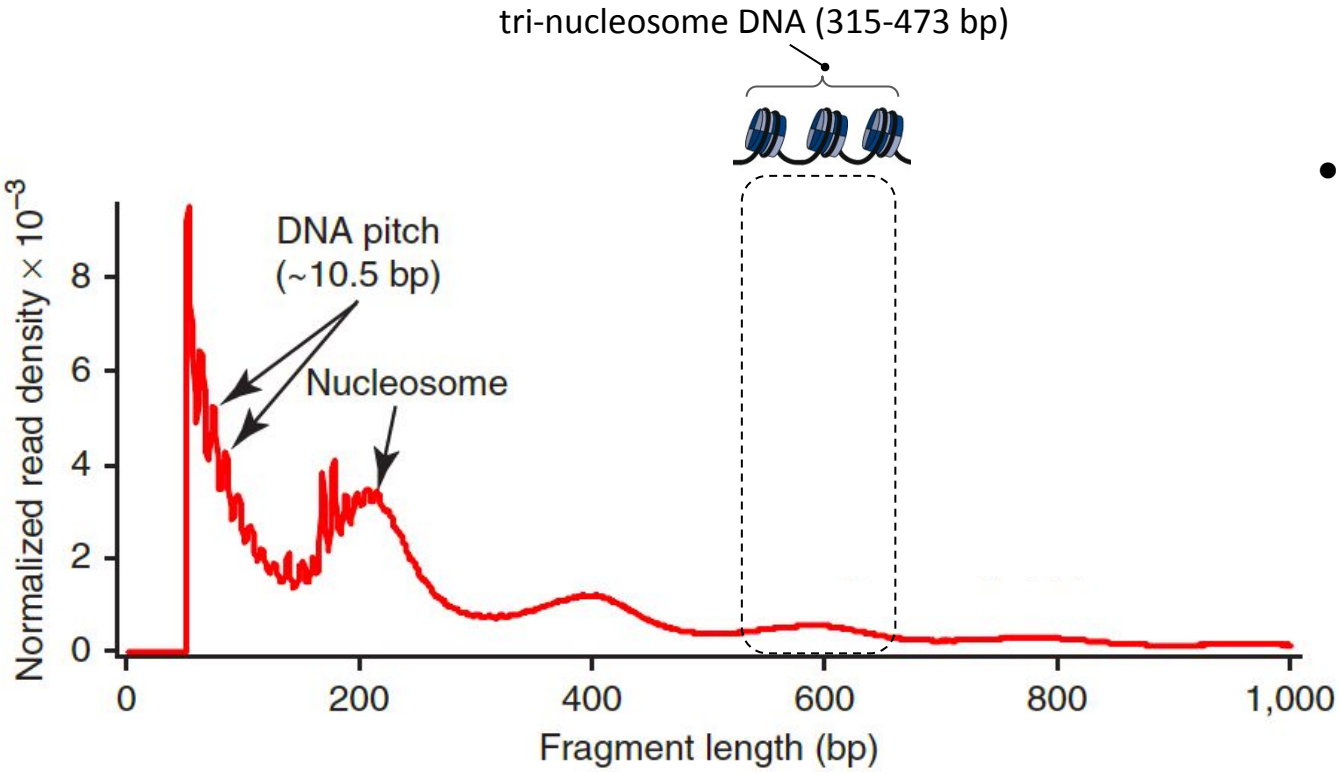
ATAC-seq proves the periodicity of nucleosome positioning



- periodicity of ~200 bp (~ nucleosomal DNA)

Buenrostro et al., 2013

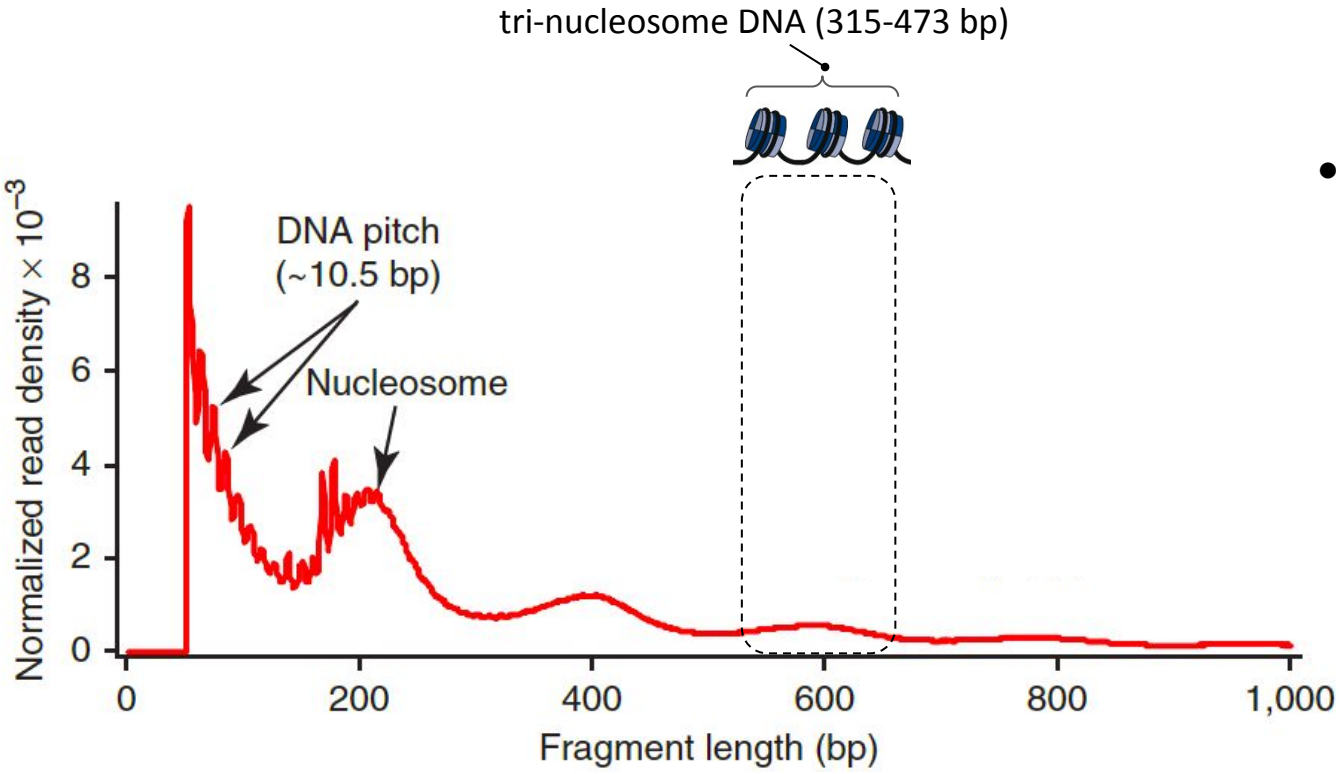
ATAC-seq proves the periodicity of nucleosome positioning



- periodicity of ~200 bp (~ nucleosomal DNA)

Buenrostro et al., 2013

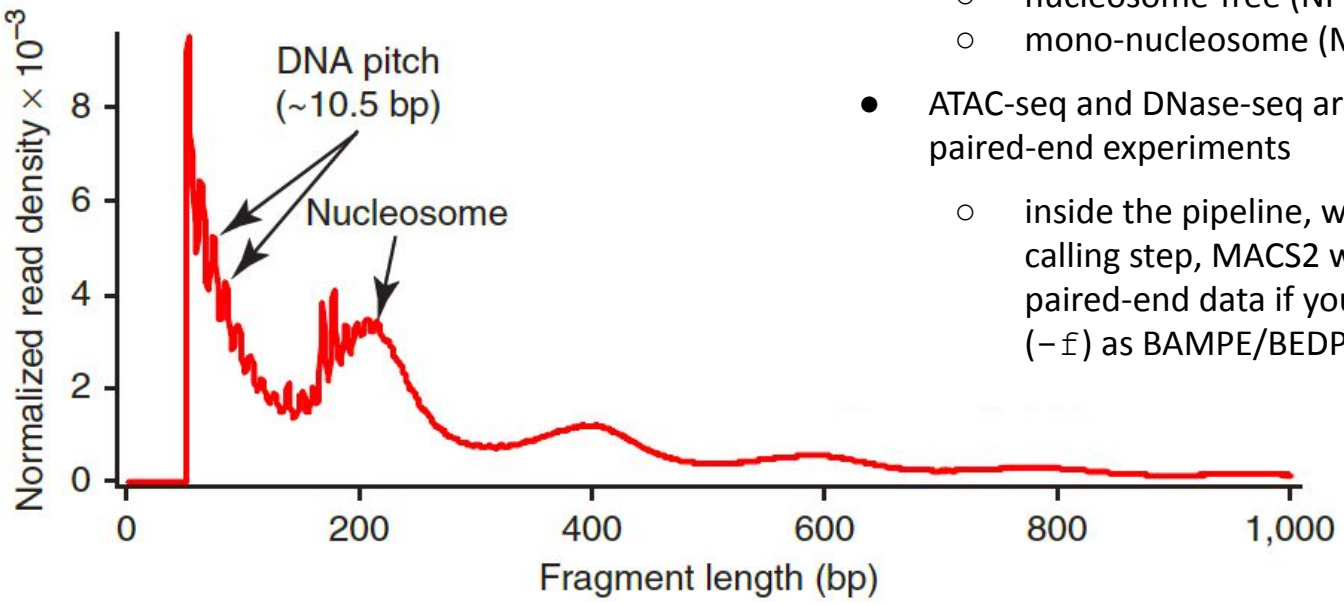
ATAC-seq proves the periodicity of nucleosome positioning



- periodicity of ~200 bp (~ nucleosomal DNA)

Buenrostro et al., 2013

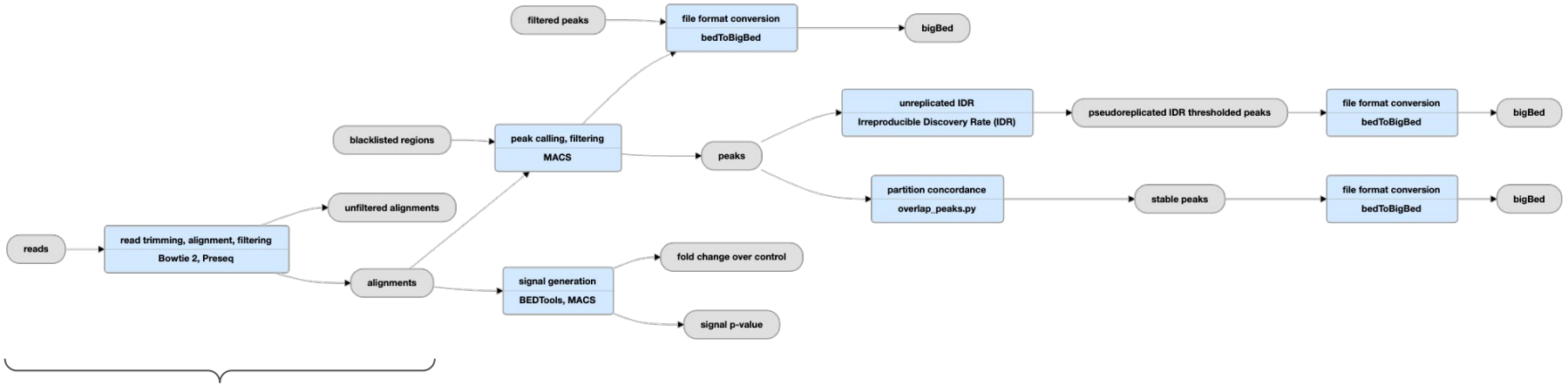
ATAC-seq proves the periodicity of nucleosome positioning



- depending on the read size you select, you can call different types of peaks
  - nucleosome-free (NF)
  - mono-nucleosome (MN)
- ATAC-seq and DNase-seq are most frequently paired-end experiments
  - inside the pipeline, when running the peak calling step, MACS2 will recognize paired-end data if you specify the format (`-f`) as BAMPE/BEDPE

Buenrostro et al., 2013

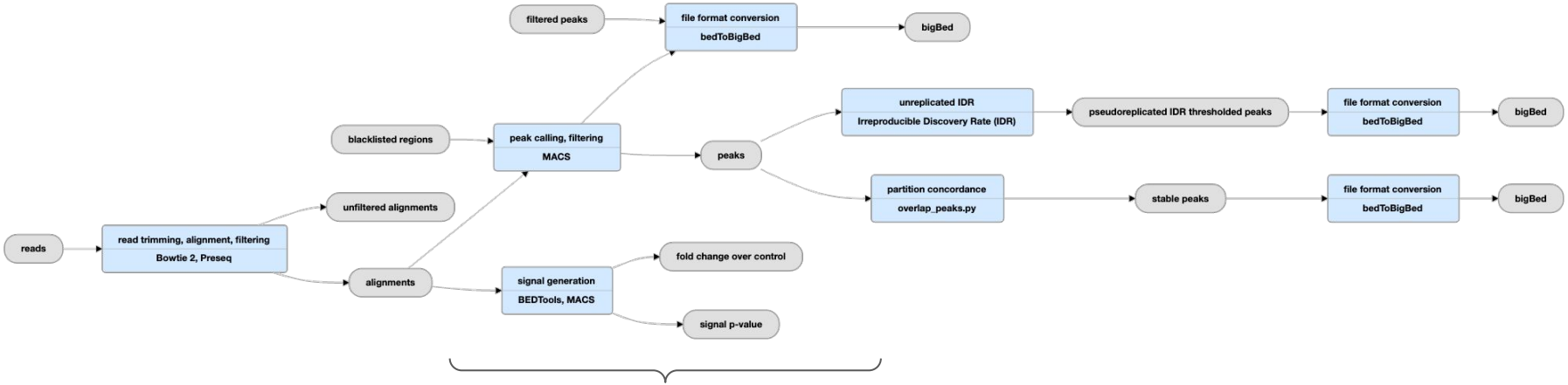
ENCODE pipeline for ATAC-seq and DNAse-seq (unreplicated experiment)



alignment step: Bowtie (for ENCODE ChIP-seq we had BWA, for *chip-nf* we had GEM)

https://github.com/ENCODE-DCC/atac-seq-pipeline

ENCODE pipeline for ATAC-seq and DNAse-seq (unreplicated experiment)



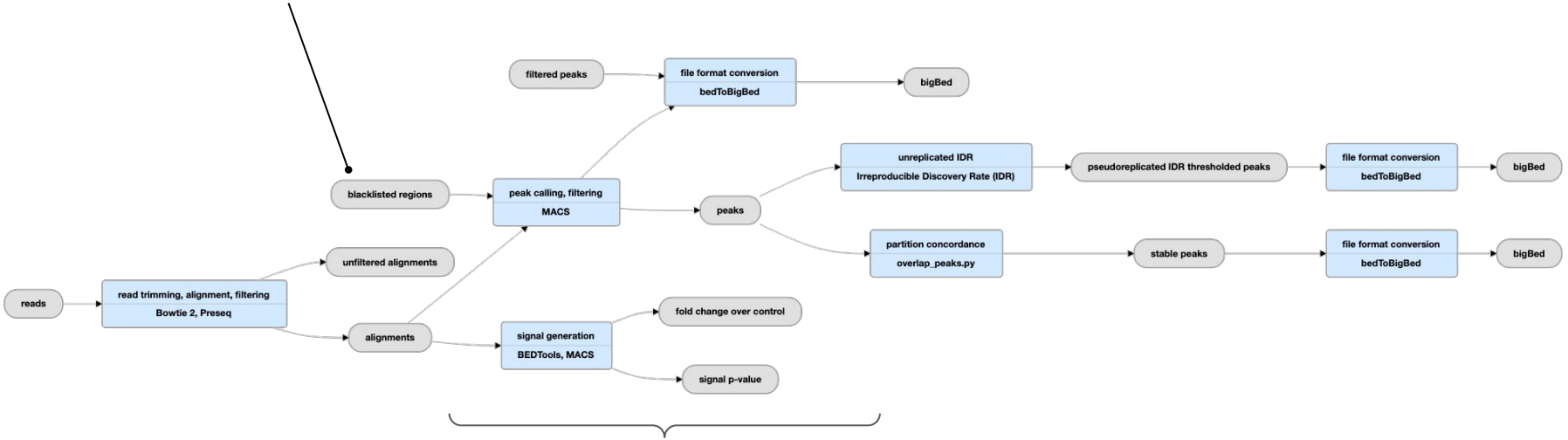
peak calling: MACS2 adapted to ATAC/DNase

- paired-end data
- no control → λ_{5k} and λ_{10k} computed on the ATAC sample

https://github.com/ENCODE-DCC/atac-seq-pipeline

ENCODE pipeline for ATAC-seq and DNAse-seq (unreplicated experiment)

blacklisted regions: regions of the genome that have anomalous, unstructured or high signal in NGS data independently of the experiment

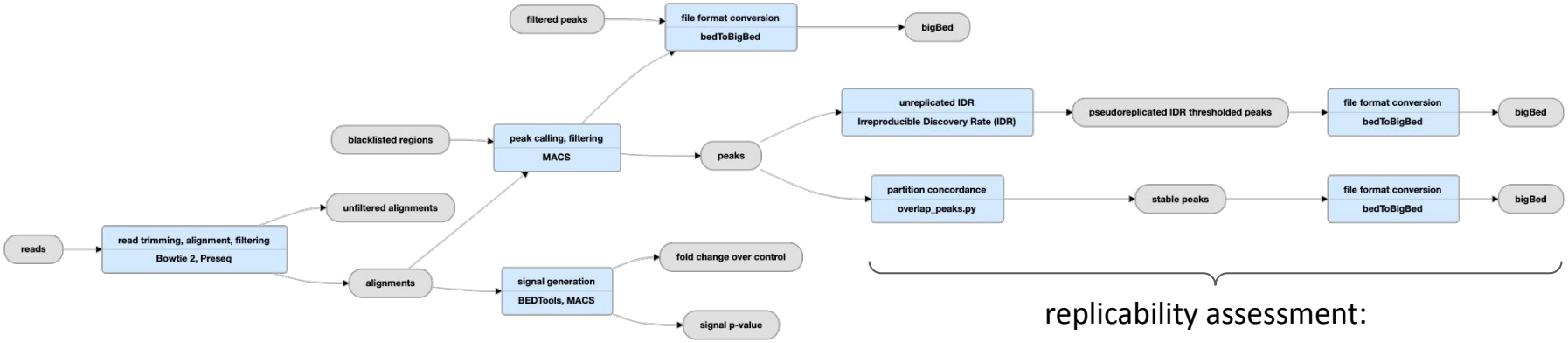


peak calling: MACS2 adapted to ATAC/DNase

- paired-end data
- no control → λ_{5k} and λ_{10k} computed on the ATAC sample

https://github.com/ENCODE-DCC/atac-seq-pipeline

ENCODE pipeline for ATAC-seq and DNAse-seq (unreplicated experiment)

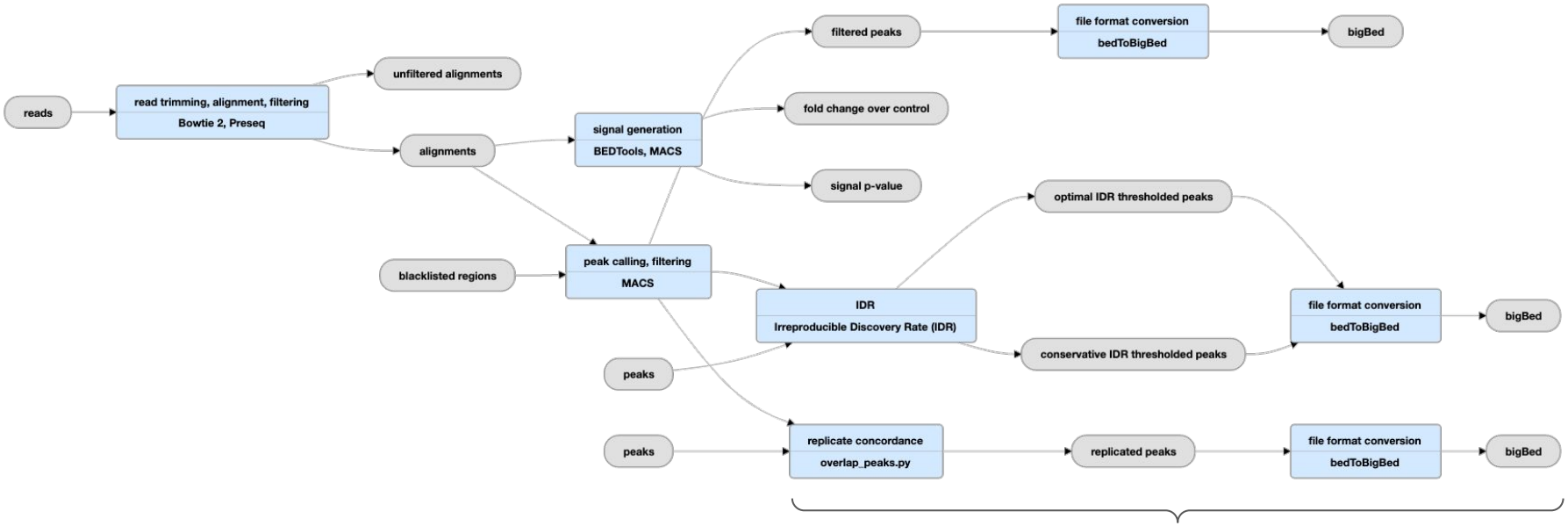


replicability assessment:

- stable peaks
- pseudoreplicated IDR thresholded peaks

https://github.com/ENCODE-DCC/atac-seq-pipeline

ENCODE pipeline for ATAC-seq and DNAse-seq (replicated experiment)



replicability assessment:

- replicated peaks
- IDR peaks: conservative or optimal IDR peaks

https://github.com/ENCODE-DCC/atac-seq-pipeline

Hands-on session

- Hands-on session 4
- Contact: beatrice.borsari@crg.eu

Hands-on sessions and references

- References:
  - Klemm et al. (2019), Nat Rev Genet
  - Crawford et al. (2006), Genome Res
  - Sabo et al. (2006), Nat Methods
  - Giresi et al. (2007), Genome Res
  - Buenrostro et al. (2013), Nat Methods
  - Corces et al. (2017), Nat Methods
  - Mieczkowski et al. (2016), Nat Commun
  - Mueller et al. (2017), Genes Dev
  - Kelly et al. (2012), Genome Res
  - Boyle et al. (2008), Cell
  - Hesselberth et al. (2009), Nat Methods
  - Furey 2012, Nat Rev Genet